**NOTAS PARA ESTADÍSTICA**

Imagina que hemos preguntado a un conjunto de *N* personas qué opinión tienen acerca de la subvención que la Comunidad Económica Europea ha concedido para la remodelación del peinado de la sardina común. Las *N* respuestas se encuentran en una escala que va de 1 a 9, donde 1 representa un total desacuerdo con la subvención, mientras que 9 quiere significar un acuerdo total.

El resultado de la medición es el siguiente:

|  |
| --- |
| 7 5 6 8 6 5 9 5 8 6 5 7 5 5 4 5 8 5 4 2 6 6 4 6 4 8 4 3 4  3 3 1 4 5 6 5 8 5 4 7 4 3 5 3 4 9 4 2  6 3 4 2 4 1 3 6 3 1 2 4  4 6 2 4 7 4 2 4 6 4 4 6 7 5 8 5 7 6 5 6 5 7 5 6      4 5 4 1 6 5 6 5 5 5 4 6 2 5 5 6 5 4 4 3 5 5 9 4 3 6 5 7 3 2     4 4 7 4 2 1 8 2 7 4 5 5 7 5 5 1 5 8 5  6 7 6 6 7 7 5 2 5 6 5 8 5 3 6 5 5 |
| Tabla 1: *Conjunto original de datos* |

Intenta responder a las siguientes preguntas:

* Cuántas personas fueron encuestadas
* Cuál fue la respuesta más frecuente
* Cuántas personas tienen, como máximo, una actitud de cuatro puntos en la escala (es decir, cuántas personas se encuentran en desacuerdo con la subvención)

Nos hemos preguntado por:

* Cuántas personas fueron encuestadas
* Cuál fue la respuesta más frecuente
* Cuántas personas tienen, como máximo, una actitud de cuatro puntos en la escala (es decir, cuántas personas se encuentran en desacuerdo con la subvención)

Es difícil responder a las tres cuestiones. Unas más que otras. ¿Cuál es el problema?

Las personas tenemos dificultades para procesar o tener en cuenta mucha información de forma simultanea. La tabla 1 muestra demasiados datos y es preciso contar con mucha paciencia y una buena vista para responder a las preguntas anteriores con seguridad.

Así pues, ¿Qué hacer? Una solución alternativa al repaso repetitivo de la tabla 1 es organizar los datos de tal forma que tengan una disposición que facilite la lectura. En este sentido, la primera acción a realizar es ordenar los datos desde el que posee el valor más pequeño hasta el que cuenta con el valor mayor.

Observa el resultado:

|  |
| --- |
| 7 5 6 8 6 5 9 5 8 6 5 7 5 5 4 5 8 5 4 2 6 6 4 6 4 8 4 3 4  3 3 1 4 5 6 5 8 5 4 7 4 3 5 3 4 9 4 2  6 3 4 2 4 1 3 6 3 1 2 4  4 6 2 4 7 4 2 4 6 4 4 6 7 5 8 5 7 6 5 6 5 7 5 6      4 5 4 1 6 5 6 5 5 5 4 6 2 5 5 6 5 4 4 3 5 5 9 4 3 6 5 7 3 2     4 4 7 4 2 1 8 2 7 4 5 5 7 5 5 1 5 8 5  6 7 6 6 7 7 5 2 5 6 5 8 5 3 6 5 5 |
| Tabla 1: *Conjunto original de datos* |

|  |
| --- |
| 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 4     4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5  5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5  5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6  6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 9 9 9 |
| Tabla 2: *Conjunto ordenado de datos* |

Observa que tiene lugar una ganancia al pasar de la tabla 1 a la tabla 2. Parece que ésta es más fácil de interpretar. No ha desaparecido ninguna información; el único cambio está en la ordenación de los mismos datos.

No obstante, la solución es parcial, puesto que aún debe ser mejorada (sigue siendo difícil responder a las preguntas).

Si observas la tabla 2, contiene una sucesión de datos con valores repetidos. Por ejemplo, el valor 1 se encuentra presente en seis ocasiones. Luego, una buena estrategia es mostrar una sola vez cada valor y hacerlo seguir por su frecuencia, es decir, por la cantidad de ocasiones en que aparece. Siguiendo este criterio, hemos conseguido la tabla 3:

|  |
| --- |
| 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 4     4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5  5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5  5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6  6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 9 9 9 |
| Tabla 2: *Conjunto ordenado de datos* |

|  |
| --- |
| 1 (6), 2 (11), 3 (12),     4 (30),  5 (40),  6 (25),  7 (14), 8 (9), 9 (3) |
| Tabla 3: *Conjunto ordenado  de "valores" y "frecuencias"* |

Aún se puede disponer la información de tal forma que resulte extremadamente fácil responder a preguntas del mismo tipo que las que hemos planteado. En la tabla 3 se ha mantenido la misma disposición que en la tabla 2. Esto es innecesario. Para disponer la información de manera óptima, vamos a generar una tabla que tenga dos columnas. En la columna primera se presentarán los valores, que representaremos con la letra X mientras que en la segunda columna se dispondrán las frecuencias, que representaremos con la letra f. Observa el resultado en la tabla 4:

|  |
| --- |
| 1 (6), 2 (11), 3 (12),     4 (30),  5 (40),  6 (25),  7 (14), 8 (9), 9 (3) |
| Tabla 3: *Conjunto ordenado  de "valores" y "frecuencias"* |

|  |  |
| --- | --- |
| X | f |
| 1 2 3 4 5 6 7 8 9 | 6 11 12 30 40 25 14 9 3 |
| Total | 150 |
| Tabla 4: *Tabla de frecuencias* | |

Como has podido leer en el título de la tabla 4, se trata de lo que llamamos, una *tabla de frecuencias*. Como ves, la estadística no se complica en exceso a la hora de bautizar sus productos. Eso está bien ¿No?

Ahora sí, la tabla de frecuencias nos permite responder a las preguntas planteadas con facilidad:

* ¿Cuántas personas fueron encuestadas?   
  Solución: 150
* ¿Cuál fue la respuesta más frecuente?   
  Solución: 5 (50 datos)
* ¿Cuántas personas tienen, como máximo, una actitud de cuatro puntos en la escala?   
  Solución: 59 (6+11+12+30)

No todas las preguntas que hemos realizado sobre el mismo conjunto de datos han exigido el mismo esfuerzo. Así, mientras que las preguntas sobre el número de datos y el valor más frecuente se han respondido con una lectura de la tabla, la tercera pregunta ha necesitado de algunas operaciones.

¿Recuerdas la pregunta?:

* Cuántas personas tienen, como máximo, una actitud de cuatro puntos en la escala   
  Solución: 59 (6+11+12+30)

Para responder a esa pregunta hemos tenido que realizar una suma: la de todas las frecuencias comprendidas entre el primer valor de la tabla y el valor que nos interesa, ambos inclusive. Esta cantidad final recibe el nombre de *frecuencia acumulada* (como puedes comprobar, la complejidad de la estadística no se encuentra en las denominaciones de sus objetos)

Muchos interrogantes requieren respuestas que se basan en las frecuencias acumuladas. Luego, es recomendable escribir esta nueva información en la tabla, de tal forma que permita respuestas directas en el futuro. Observa el resultado:

|  |  |
| --- | --- |
| X | f |
| 1 2 3 4 5 6 7 8 9 | 6 11 12 30 40 25 14 9 3 |
| Total | 150 |
| Tabla 4: *Tabla de frecuencias* | |

|  |  |  |
| --- | --- | --- |
| X | f | F |
| 1 2 3 4 5 6 7 8 9 | 6 11 12 30 40 25 14 9 3 | 6 17 29 59 99 124 138 147 150 |
| Total | 150 |  |
| Tabla 5: Tabla de frecuencias de tres columnas | | |

Imagina ahora que hemos preguntado a 25 personas por su provincia de nacimiento, obteniendo los siguientes resultados:

|  |
| --- |
| Sevilla, Córdoba, Huelva, Granada, Granada,  Cádiz, Huelva, Almería, Sevilla, Málaga,  Jaén, Huelva, Almería, Sevilla, Córdoba,  Almería, Málaga, Córdoba, Málaga, Almería,  Málaga, Sevilla, Cádiz, Jaén, Cádiz |

Construye una tabla de frecuencias con la información sobre las provincias de nacimiento, utilizando la siguiente equivalencia:

|  |  |
| --- | --- |
| Provincia | Código |
| Almería  Cádiz  Córdoba  Granada  Huelva  Jaén  Málaga  Sevilla | 1 2 3 4 5 6 7 8 |

Quizá tu respuesta haya sido ésta:

|  |  |  |  |
| --- | --- | --- | --- |
| Provincia | Código | f | F |
| Almería  Cádiz  Córdoba  Granada  Huelva  Jaén  Málaga  Sevilla | 1 2 3 4 5 6 7 8 | 4 3 2 2 3 2 4 4 | 4 7 10 12 15 17 21 25 |
| Tabla 6: Distribución por provincias | | | |

Si es así ... ¡Error!

La estadística es como un cuchillo. Constituye una herramienta extraordinariamente útil para multitud de menesteres cotidianos. Pero también puede utilizarse para fines, digamos, inapropiadamente peligrosos. El cuchillo ignora qué se está haciendo con él. La responsabilidad es de quien lo empuña.

¿Qué sentido tiene acumular frecuencias en el problema que se ha planteado sobre las provincias?. Por ejemplo, ¿Qué significado tiene la cantidad 12 que acompaña al valor 4 (Granada)? Sólo se puede hacer una lectura: hay doce personas que *han nacido en Granada o en menos que Granada*. No podemos afirmar que Córdoba, Cádiz o Almería sean *menos provincia de nacimiento* que Granada.

La diferencia esencial entre el problema de las provincias de nacimiento y el de las respuestas a la escala de acuerdo, se encuentra en el tipo de variable. En el caso de las provincias, éstas *no* pueden ordenarse en función de ser más o ser menos "provincia de nacimiento" (se pueden ordenar según número de habitantes, extensión, altitud media, etc. Pero no en función de ser más o ser menos provincia de nacimiento).

Luego, la acumulación de frecuencias sólo procede si los valores de la variable que se está estudiando se pueden ordenar. Así, la respuesta correcta al problema debe ser:

|  |  |  |
| --- | --- | --- |
| Provincia | Código | f |
| Almería  Cádiz  Córdoba  Granada  Huelva  Jaén  Málaga  Sevilla | 1 2 3 4 5 6 7 8 | 4 3 2 2 3 2 4 4 |
| Tabla 7: Distribución por provincias | | |

Retomamos ahora el problema de las actitudes frente a la subvención europea.

La tabla de frecuencias no termina donde la hemos dejado. Se puede añadir más información útil en la que basar respuestas para otras preguntas.

Por ejemplo ¿Cuántas personas han respondido con una actitud media (valor 5)? Solución: 40. Observa ahora la siguiente tabla y responde a la misma pregunta.

|  |  |
| --- | --- |
| X | f |
| 1 2 3 4 5 6 7 8 9 | 200 170 120 60 40 60 120 170 200 |
| Total | 1140 |
| Tabla 7: *Nueva tabla de frecuencias* | |

¿Qué ocurre ahora?

En la tabla 7 ha cambiado el conjunto de datos. Ahora son 1140, frente a los 150 del colectivo anterior. Una misma frecuencia, en este caso f=40, no tiene la misma interpretación en ambas tablas. ¿Qué ha cambiado?: la importancia relativa de la frecuencia, puesto que f=40 frente a N=150 es diferente a f=40 frente a N=1140. De hecho, el valor 5 pasa incluso de ser el más frecuente al menos presente.

La solución se encuentra en expresar las frecuencias en términos relativos en vez de absolutos. Esto es precisamente lo que consiguen las proporciones: expresar una cantidad con respecto al total. Así, añadimos una nueva columna, conteniendo las frecuencias relativas (fr) que surgen de hacer la operación *fr = f / N*. Observa el resultado comparando el obtenido con cada una de las dos tablas afectadas en este problema (4 y 7):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Nuevos datos | |  | Datos anteriores | |
| X |  | f | fr |  | f | fr |
| 1 2 3 4 5 6 7 8 9 |  | 200 170 120 60 40 60 120 170 200 | 0,1754 0,1491 0,1053 0,0526 0,0351 0,0526 0,1053 0,1491 0,1754 |  | 6 11 12 30 40 25 14 9 3 | 0,0400 0,0733 0,0800 0,2000 0,2667 0,1667 0,0933 0,0600 0,0200 |
| Total |  | 1140 | 1,0000 |  | 150 | 1,0000 |
| Tabla 8: *Comparación entre dos tablas de frecuencias* | | | | | | |

Observa que el valor 5 pasa de contar con una frecuencia relativa fr=0,2667 (más de la cuarta parte) a fr=0,0351 al ser comparado, respectivamente, con un total de n=150 a n=1140.

Un aspecto de interés se encuentra en la fila de los totales. Observa que el resultado es 1,0000 en los dos casos. Esto debe ocurrir siempre. Lo que se hacer al traducir las frecuencias absolutas a las relativas es *unificar* el referente. En el conjunto de datos de la tabla 4, el referente absoluto es 150. En el conjunto de datos de la tabla 7, el referente absoluto es 1140. No podemos comparar frecuencias de conjuntos de datos diferentes porque los referentes son diferentes. Para que la comparación sea factible es necesario *unificar*. Dado que las proporciones se expresan en *tantos por uno*, es posible comparar frecuencias entre tablas. En otros términos: para interpretar una frecuencia absoluta necesitamos conocer el número total de datos puesto que, según hemos visto, el número de datos condiciona la importancia de una frecuencia. Pero para interpretar una frecuencia relativa expresada como una proporción no es necesario conocer el número total de datos, puesto que aquí el referente es constante de una tabla a otra: 1,0000.

Sin embargo, no se terminó el proceso de enriquecimiento de la tabla.

Las proporciones se expresan siempre en cantidades que se situan entre 0 y 1. Es decir, las proporciones son números decimales. Y las personas también nos sentimos incómodas con las cantidades decimales. ¿Solución?

Por lo general, cuando se exponen los resultados de una encuesta en un medio de comunicación, lo habitual es utilizar otro tipo de frecuencias relativas: los porcentajes.

El principio que rige la utilización de los porcentajes es el mismo que para las proporciones: utilizar un referente fijo de tal forma que no sea necesario contar con el número total de datos para interpretar una frecuencia. La diferencia entre los porcentajes y las proporciones es que los primeros utilizan el referente 100, mientras que las proporciones utilizan el 1.

Luego, conseguir los porcentajes es muy fácil si se cuenta con las proporciones: bastará con multiplicar a éstas por 100:

|  |  |  |  |
| --- | --- | --- | --- |
| X | f | fr | % |
| 1 2 3 4 5 6 7 8 9 | 6 11 12 30 40 25 14 9 3 | 0,0400 0,0733 0,0800 0,2000 0,2667 0,1667 0,0933 0,0600 0,0200 | 4,00 7,33 8,00 20,00 26,67 16,67 9,33 6,00 2,00 |
| Total | 150 | 1,0000 | 100,00 |
| Tabla 9: *Tabla de frecuencias con porcentajes* | | | |

Podemos completar también la tabla que se refiere a las provincias de nacimiento:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Provincia | Código | f | fr | % |
| Almería  Cádiz  Córdoba  Granada  Huelva  Jaén  Málaga  Sevilla | 1 2 3 4 5 6 7 8 | 4 3 2 2 3 2 4 4 | 0,16 0,12 0,08 0,08 0,12 0,08 0,16 0,16 | 16 12 8 8 12 8 16 16 |
| Tabla 10: Distribución por provincias | | | | |

Y, por último...

Como sabemos, la variable *actitud frente a la subvención* admite orden entre sus valores. Luego, para completar la tabla, bastará con acumular sus frecuencias:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | f | fr | % | F | Fr | %a |
| 1 2 3 4 5 6 7 8 9 | 6 11 12 30 40 25 14 9 3 | 0,0400 0,0733 0,0800 0,2000 0,2667 0,1667 0,0933 0,0600 0,0200 | 4,00 7,33 8,00 20,00 26,67 16,67 9,33 6,00 2,00 | 6 17 29 59 99 124 138 147 150 | 0,0400 0,1133 0,1933 0,3933 0,6600 0,8267 0,9200 0,9800 1,0000 | 4,00 11,33 19,33 39,33 66,00 82,67 92,00 98,00 100,00 |
| Total | 150 | 1,0000 | 100,00 |  |  |  |
| Tabla 11: *Tabla de frecuencias completa* | | | | | | |